



# Clinical usefulness of gene-expression profile to rule out acute rejection after heart transplantation: CARGO II

**Maria G. Crespo-Leiro<sup>1†</sup>, Jörg Stypmann<sup>2†</sup>, Uwe Schulz<sup>3</sup>, Andreas Zuckermann<sup>4</sup>, Paul Mohacs<sup>5</sup>, Christoph Bara<sup>6</sup>, Heather Ross<sup>7</sup>, Jayan Parameshwar<sup>8</sup>, Michal Zakliczyński<sup>9</sup>, Roberto Fiocchi<sup>10</sup>, Daniel Hoefer<sup>11</sup>, Monica Colvin<sup>12</sup>, Mario C. Deng<sup>13</sup>, Pascal Leprince<sup>14</sup>, Barbara Elashoff<sup>15</sup>, James P. Yee<sup>15</sup>, and Johan Vanhaecke<sup>16\*</sup>**

<sup>1</sup>Instituto de Investigación Biomédica de A Coruña (INIBIC), Complejo Hospitalario Universitario de A Coruña (CHUAC), SERGAS, Universidade da Coruña (UDC), A Coruña, Spain; <sup>2</sup>Department of Cardiovascular Medicine, University Hospital Muenster, Muenster, Germany; <sup>3</sup>Ruhr University of Bochum, Bad Oeynhausen, Germany; <sup>4</sup>Medical University of Vienna, Vienna, Austria; <sup>5</sup>University Hospital Bern, Bern, Switzerland; <sup>6</sup>Hannover Medical School, Hannover, Germany; <sup>7</sup>Toronto General Hospital, Toronto, Canada; <sup>8</sup>Papworth Hospital, Papworth Everard, Cambridge, UK; <sup>9</sup>Silesian Center for Heart Disease, Zabrze, Poland; <sup>10</sup>Ospedali Riuniti di Bergamo, Bergamo, Italy; <sup>11</sup>Innsbruck Medical University, Innsbruck, Austria; <sup>12</sup>University of Minnesota, Minneapolis, MN, USA; <sup>13</sup>David Geffen School of Medicine, University of California, Los Angeles, CA, USA; <sup>14</sup>Groupe Hospitalier Pitié-Salpêtrière, Paris, France; <sup>15</sup>CareDx, Brisbane, CA, USA; and <sup>16</sup>Department of Cardiology, University Hospital Gasthuisberg, Leuven, Belgium

Received 4 May 2015; revised 25 September 2015; accepted 25 November 2015

## Aims

A non-invasive gene-expression profiling (GEP) test for rejection surveillance of heart transplant recipients originated in the USA. A European-based study, Cardiac Allograft Rejection Gene Expression Observational II Study (CARGO II), was conducted to further clinically validate the GEP test performance.

## Methods and results

Blood samples for GEP testing (AlloMap<sup>®</sup>, CareDx, Brisbane, CA, USA) were collected during post-transplant surveillance. The reference standard for rejection status was based on histopathology grading of tissue from endomyocardial biopsy. The area under the receiver operating characteristic curve (AUC-ROC), negative (NPVs), and positive predictive values (PPVs) for the GEP scores (range 0–39) were computed. Considering the GEP score of 34 as a cut-off (>6 months post-transplantation), 95.5% (381/399) of GEP tests were true negatives, 4.5% (18/399) were false negatives, 10.2% (6/59) were true positives, and 89.8% (53/59) were false positives. Based on 938 paired biopsies, the GEP test score AUC-ROC for distinguishing  $\geq 3A$  rejection was 0.70 and 0.69 for  $\geq 2-6$  and  $>6$  months post-transplantation, respectively. Depending on the chosen threshold score, the NPV and PPV range from 98.1 to 100% and 2.0 to 4.7%, respectively.

## Conclusion

For  $\geq 2-6$  and  $>6$  months post-transplantation, CARGO II GEP score performance (AUC-ROC = 0.70 and 0.69) is similar to the CARGO study results (AUC-ROC = 0.71 and 0.67). The low prevalence of ACR contributes to the high NPV and limited PPV of GEP testing. The choice of threshold score for practical use of GEP testing should consider overall clinical assessment of the patient's baseline risk for rejection.

## Keywords

Heart transplant • Organ rejection • Tests • Molecular diagnostics • AlloMap • Rejection surveillance

## Introduction

A non-invasive gene-expression profiling (GEP) test (AlloMap<sup>®</sup>, CareDx, Brisbane, CA, USA) was first demonstrated to discriminate

acute cellular rejection from quiescence in cardiac transplant recipients in a US-based clinical study, Cardiac Allograft Rejection Gene Expression Observational Study (CARGO).<sup>1,2</sup> Estimates of the negative predictive value (NPV) associated with a GEP score

<sup>†</sup>These authors contributed equally to this work and shared co-first authorship.

\*Corresponding author. Tel: +32 16 3 44251, Fax: +32 16 3 44240, Email: [johan.vanhaecke@uzleuven.be](mailto:johan.vanhaecke@uzleuven.be)

© The Author 2016. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

below a nominal threshold level (e.g. 34) were refined during re-analyses of the CARGO dataset during the US Food and Drug Administration (US FDA) regulatory clearance of AlloMap as a *de novo* 510 k *in vitro* diagnostic test (see Supplementary material online, Appendix S1). The randomized controlled Invasive Monitoring Attenuation through Gene Expression (IMAGE) study showed the non-inferiority of clinical outcomes in patients managed with GEP for rejection surveillance compared with patients managed with conventional biopsy surveillance.<sup>3</sup> The International Society of Heart and Lung Transplantation (ISHLT) guidelines for care of heart transplant recipients recommend that the GEP test can be used as a non-invasive method for ruling out moderate to severe acute cellular rejection in asymptomatic patients (i.e. without clinical suspicion of allograft dysfunction).<sup>4</sup> The CARGO II study was designed to evaluate GEP in a different population from CARGO. There are very few published clinical validation studies for commercially available diagnostic tests based on gene-expression profiling of peripheral blood, and AlloMap is the only product of this kind for use in heart transplantation.<sup>5</sup>

Therefore, the objective of the current report was to validate the clinical performance of the gene-expression profiling technology in an entirely independent large study population. The hypothesis of this study is that the GEP test can discriminate acute cellular rejection from non-rejection status, with high negative predictive values and overall accuracy similar to that previously demonstrated in CARGO. Antibody-mediated rejection cases were not included in these computations. The GEP test performance was evaluated in two post-transplantation time windows:  $\geq 2$ –6 months and  $> 6$  months, because prior evidence had indicated that the risk of rejection and performance of the GEP test may both be sensitive to time post-transplantation.<sup>1,6,7</sup>

## Methods

### CARGO II study design

The CARGO II was a prospective, observational, multi-centre study with blood samples and associated clinical data collected from follow-up visits of cardiac transplantation recipients from May 2005 through February 2009 (ClinicalTrials.gov Identifier: NCT00761787). The 17 participating centres (13 Western European and 4 North American) obtained Institutional Review Board approvals and signed written informed consents from participating patients. The study was conducted in accordance with the Good Clinical Practices and Declaration of Helsinki.

Venous blood was collected at clinic visits during which heart transplant recipients underwent long-term ongoing rejection surveillance, with or without endomyocardial biopsy (EMB). The blood samples were collected before EMB. In general, patients were asked not to drink or eat anything within 8 h before the procedure. An outline of the clinical information collected at each visit, including patient inclusion and exclusion criteria, and sample collection procedures, is provided in Supplementary material online, Appendix S2.

### Reference standard for rejection status

The reference standard for rejection status was the histology-based rejection grade (using the ISHLT 1990 grades: 0, 1A, 1B, 2, 3A, 3B, or 4) for heart tissue obtained from an EMB performed on the same day as the collection of blood for the GEP test.<sup>8</sup> In 2005, the grading system was revised as follows: Grade 0R-no rejection (no change from the 1990

grading system), Grade 1R-mild rejection (1990 Grades 1A, 1B, and 2), Grade 2R-moderate rejection (1990 Grade 3A), and Grade 3R-severe rejection (1990 Grades 3B and 4).<sup>9</sup> In routine clinical practice, patient care decisions are most strongly influenced by findings of rejection grades  $\geq 3A$  (2R), which usually lead to a treatment intervention. The histopathology grade is rendered by the local pathologist, who has viewed the histology slides with light microscopy. In the CARGO II study, a subset (see below) of all locally graded biopsy slides was also sent for review by three other pathologists. The intention of this independent assignment of rejection grades by a central panel of pathologists was to enhance the accuracy of the reference standard for rejection that was used to judge the performance of the GEP test.<sup>10,11</sup>

Because of the high proportion of local EMB scores with Grade 0, and the limited volume capacity for independent scoring by the reviewing central pathologists, only one local rejection Grade 0 sample from each subject was sent for central pathology reading; if a subject had more than one sample with a local biopsy grade of 0, then one of the samples was chosen at random. At most, only one Grade 1A sample from each subject was sent for central pathology reading; if a subject had more than one sample with a local biopsy grade of 1A, then one of the samples was chosen at random. All samples that had local biopsy grades of 1B, 2, and 3A or higher were sent for the independent central pathology grading. For this report, each rejection case was qualified for use as a reference standard of rejection if a moderate or more severe grade of rejection, [i.e. ISHLT 1990 grade of  $\geq 3A$  (2R)] was reported by the local pathologist and this grade  $\geq 3A$  (2R) was independently confirmed by at least one of three central reviewing pathologists. All other centrally reviewed biopsy slides were classified as quiescent (i.e. Grade 0 assigned by the local pathologist and at least 2 of 3 central pathologists) or other non-rejection cases (i.e. did not meet either the criteria for classification as quiescent or rejection). All samples that had central pathology results and GEP score results were included in the analyses.

### Methodology of gene-expression profiling

For the GEP test, a whole blood sample was collected into a BD Vacutainer® CPT™ Cell Preparation tube (Becton Dickinson, NJ, USA) with sodium citrate anticoagulant (CPT tube). Peripheral blood mononuclear cells were isolated, lysed by RLT buffer (Qiagen) and the released ribonucleic acid (RNA) was stabilized and frozen. Following shipment to the CareDx reference laboratory in Brisbane, CA, USA, RNA was further purified from the lysate; finally, complementary deoxyribonucleic acid (cDNA) was generated for use in the GEP test. The GEP test was conducted in the CareDx Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory. The test uses real-time polymerase chain reaction (RT-PCR) to quantitate expression levels of a preselected panel of 20 genes: 11 genes informative about allograft rejection and 9 genes for normalization and quality control, used in the calculation of a GEP logarithmic score that ranges from 0 to 39, Supplementary material online, Appendix S3.<sup>1</sup>

### Sample selection

The blood samples included in this study conformed to the following inclusion criteria: obtained at least 55 days post-transplantation;  $> 30$  days after transfusion of blood products;  $> 21$  days after administration of  $\geq 20$  mg/day of prednisone; and  $> 60$  days after treating a prior rejection.

### Statistical analysis

To calculate the GEP score sensitivity (ability to detect acute cellular rejection) and specificity (ability to detect the absence of rejection) and generate an associated area under the receiver operating characteristic

(AUC-ROC) curve, positive predictive value (PPV), and NPV, each GEP sample was correlated to a paired, histology-based rejection grade (using the ISHLT 1990 grades: 0, 1A, 1B, 2, 3A, 3B, or 4 or equivalent ISHLT 2005 grades) for heart tissue obtained from an EMB performed on the same day.<sup>8,9</sup>

The main metric for validating the GEP test performance was the AUC-ROC, which is a widely used metric for new screening test discrimination.<sup>12,13</sup> In early 2008, during the regulatory clearance of AlloMap as a FDA 510 k *in vitro* diagnostic test, the GEP test performance characteristics from the original CARGO study were reanalysed in cooperation with reviewing statisticians from the US FDA. The main change in the FDA-guided analyses compared with the methods used in the Deng 2006 publication was that all mild (i.e. <moderate [grade 3A (2R)]) rejection cases were included in the FDA-guided analyses which thereby include the full intended use population. Incorporating these mild rejection cases is called an intent-to-diagnosis (ITD) analysis (the analog of intent-to-treat or ITT analysis for randomized clinical trials of therapeutics). All centrally graded samples including all non-rejection samples (i.e. both quiescent and other samples that did not qualify to be classified as rejection) were included in the analysis; Supplementary material online, Appendix S1. The current CARGO II analyses use the same FDA-recommended methodology that was used in the analyses of the CARGO data: AUC-ROC of GEP score was estimated using a non-parametric method proposed by Obuchowski,<sup>14</sup> altered slightly to account for the multiple records per subject.

The AUC-ROC and 95% confidence intervals were calculated using a bootstrap method for the data overall and also by time period:  $\geq 2$ –6 months ( $\geq 55$ –182 days) post-transplantation and  $> 6$  months ( $> 182$  days) post-transplantation.

Clinical performance metrics were further characterized by computing the NPV and PPV for acute cellular rejection for each GEP ordinal score from 12 to 39. These characteristics were computed for two time periods post-transplantation:  $\geq 2$ –6 months and  $> 6$  months.

The calculation of NPV and PPV involved an adjustment for prevalence of rejection, since not all local Grades 0 and 1A biopsies were sent for central pathology, as described. The adjustment was applied separately to the different time periods post-transplantation. The adjusted prevalence calculation method is shown in the statistical appendix, Supplementary material online, Appendix S4.

Logistic mixed-effects models were used to predict rejection status from GEP score, both in the absence and in the presence of other clinical variables (age, gender, primary indication for transplant, cytomegalovirus status, and pre-transplant use of a left ventricular assist device).

## Minimization of bias

The study used a prospective design in which the blood samples and associated heart biopsy slides were selected from patients enrolled in the CARGO II Study. Sample labels were encrypted and the lab technicians were blinded to the patient data associated with each sample, minimizing any operator bias in determination of the GEP test scores.

Biopsy slides were de-identified prior to centralized pathology review and grading. Since patients may have more than one sample, the method of Emir *et al.*<sup>15</sup> was used to analyse the data. This method assigns a weight to each sample based on the number of samples from the same patient within a category (rejection or non-rejection).

## Results

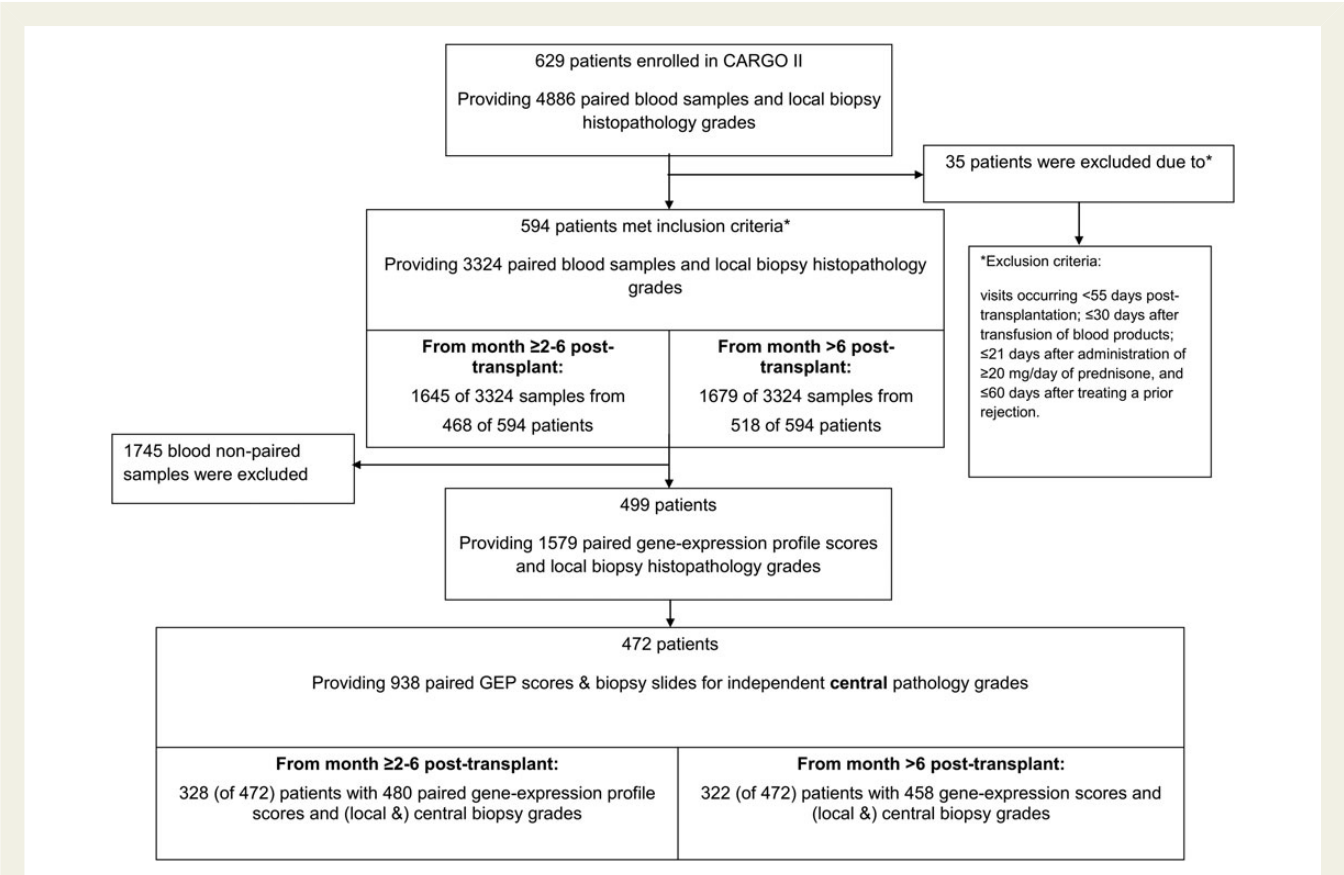
From the total CARGO II enrollment, 499 patients had 1579 visits with paired EMB histopathology rejection grades and GEP scores

that met inclusion criteria for the study analyses (Figure 1). Patients were predominantly Caucasian (89%), male (79%), and 50.8 years old on average, at the time of transplantation (Table 1). The first study visit occurred prior to 6 months post-transplantation in 88% of the patients. Moderate to severe acute cellular rejection was reported by the local pathologists in 53/1645 (3.2%) of biopsies at  $\geq 2$ –6 months post-transplantation [representing 47/468 (10.0%) of patients] and in 53/1679 (3.2%) of biopsies at  $> 6$  months post-transplantation [representing 41/518 (7.9%) of patients]. The prevalence of moderate to severe grade rejection was 106/3325 (3.2%) in the population as a whole. When the rejection was also required to be confirmed by at least one central pathologist [(i.e. rejection was defined as 1 local grade  $\geq 3A$  (2R) plus  $\geq 1$  central grade  $\geq 3A$  (2R))], the prevalence was reduced to 2.0%.

The patient population with histopathology slides selected for independent central panel pathology rejection grading included 328 patients with 480 GEP test scores at  $\geq 2$ –6 months post-transplantation, while 322 patients provided 458 GEP test scores at  $> 6$  months post-transplantation. Of a total of 938 centrally graded biopsies, 46 (4.9%) were deemed rejections, whereas 246 (26.2%) were classified as quiescent. The remaining 646 (68.9%) biopsies were intermediate (i.e. did not meet the criteria to qualify as either a rejection or a quiescent reference standard criteria). The implications of this subset with discordant grades from the pathologists are addressed in the discussion section of this paper. Of the 938 biopsies sent for central review, 919 were surveillance biopsies, 9 were for suspected rejection, and 10 were performed for other reasons. Out of the 46 centrally confirmed rejections, 43 were from surveillance biopsies, and 3 were biopsies performed because of suspected rejection. The three clinically suspected rejections that were confirmed by biopsy had GEP scores of 29, 34, and 37. The average GEP score of samples taken at the time of a biopsy-proven  $\geq 3A$  (2R) rejection was significantly higher (28.3, SD 5.3) than the average GEP score of all samples taken at times when the biopsy revealed a non-rejection status (25.6, SD 5.9),  $P = 0.001$ . Of the 938 centrally graded biopsies, the distribution of local grades was as follows (*n*, %): Grade 0 (384, 40.9%); Grade 1A (245, 26.1%); Grade 1B (68, 7.2%); Grade 2 (131, 14.0%); Grade 3A (70, 7.5%); Grade 3B (5, 1%); Grade 4 (1, <1%); local grades were not available for 34 (3.6%) biopsies. Of the 109 biopsies that were graded  $\geq 3A$  (2R) by at least one central reviewer, 63 (58%) received a local grade of 2 or less (Table 2).

The mean AUC-ROC for the GEP test scores in the  $\geq 2$ –6 months post-transplantation period was 0.70 with 95% confidence interval from 0.67 to 0.73. The mean AUC-ROC for the GEP test scores for the  $> 6$  months post-transplantation period was 0.69 with 95% confidence interval from 0.66 to 0.72. The GEP scores by time post-transplant are shown in Figure 2.

The CARGO II-based GEP NPV and PPV performance characteristics are shown in Tables 3 and 4. A GEP test score of  $\geq 34$  (often used in clinical practice as a threshold score for patients who are  $> 6$  months post-transplantation) corresponded to histology-based grade  $\geq 3A$  (2R) rejection with a PPV of 4.0% at months  $\geq 2$ –6 post-transplantation, and 4.3% at  $> 6$  months post-transplantation. The NPVs were 98.4% at months  $\geq 2$ –6 post-transplantation and 98.3% at  $> 6$  months post-transplantation. In both time windows, the NPVs increased from 98.3 to  $> 99.0\%$  for decreasing threshold



**Figure 1** Source of patients and samples included in the area under the receiver operating characteristic curve calculations and modelling of negative predictive value (NPV) and positive predictive value (PPV) performance of gene-expression profiling in the study population. Local biopsy (and pathology grading): local centre's pathologist's grade of rejection of an endomyocardial biopsy and a paired blood sample for gene-expression profiling available for the patient visit. Central pathology review: a central panel of pathologists provided independent grading of rejection of the endomyocardial biopsy slides and a paired blood sample for gene-expression profiling available for the patient visit. Note: the negative predictive value and positive predictive value calculations used the local biopsy findings to adjust for prevalence rates for rejection. The central biopsies were used to ascertain the reference standards for Grade 3A or more severe rejection grades vs. lesser grades of rejection.

values below 34. The corresponding PPVs decreased from 4.3 to 2.1. The number (and %) of patients with centrally adjudicated  $\geq 3R$  (2R) rejection below or above GEP score of 34 is presented in Table 5. Using the score of 34 as a cut-off for patients who are  $>6$  months post-transplantation, 95.5% (381/399) of GEP tests were true negatives, 4.5% (18/399) were false negatives, 10.2% (6/59) were true positives, and 89.8% (53/59) were false positives.

The probability of rejection using a threshold of  $<34$  as a negative test moves from 3.2% pretest to 1.7% post-test, i.e. the NPV goes from 96.8% without the GEP test to 98.3–98.4% with a negative test, which occurred in 82.5–85% of the population. For  $\sim 15\%$  of positive tests, the patients may be considered for a follow-up biopsy, 96% of which will be negative.

In the univariate model, GEP score and age were significantly associated with  $\geq 3A$  (2R) acute cellular rejection (Table 6). In the multivariate model, after adjustment for clinical variables, a decreasing GEP score ( $P = 0.004$ ) and age  $\geq 54$  years ( $P = 0.008$ ) were associated with decreased probability of  $\geq 3A$  (2R) rejection. Subgroup analyses of clinical variables for association of GEP score and  $\geq 3A$  (2R) acute cellular rejection are shown in Table 7.

## Discussion

The performance characteristics of GEP scores to discriminate acute cellular rejection from non-rejection status of heart transplant recipients in CARGO II are very similar to the performance of GEP scores reported from CARGO, based on similar FDA-recommended analyses applied to these two independent data sets.<sup>6</sup> This additional independent clinical validation is important, as such confirmatory studies are rare in the field of molecular diagnostic tests based on gene-expression profiling technologies; the hypothesis that the AlloMap GEP signature generated from assay of mRNA purified from circulating peripheral blood cells is informative of the immune status of the allograft is thus further validated by the results in this report.

The CARGO II study population is similar to most heart transplant study cohorts in demographic composition,<sup>16</sup> but has a higher percentage of Caucasians than the US-based CARGO study. The 3.2% prevalence of moderate to severe rejection found in locally graded biopsies and estimated 13% incidence of rejection in patients is lower than the 66% rejection rate in patients reported during the

**Table 1** Baseline characteristics of study patients

Characteristics	n = 499 <sup>a</sup>	n = 41 rejections	n = 431 non-rejection	P-values
Age, years, mean (standard deviation)	50.8 (13.7)	46.1 (13.9)	51.2 (13.5)	0.011
Male sex	393 (79)	34 (83)	338 (78)	0.689
Race				
Caucasian	443 (89)	35 (85)	386 (90)	0.427
African American	16 (3)	3 (7)	13 (3)	0.154
Asian	11 (2)	1 (2)	8 (2)	0.562
Hispanic	9 (2)	1 (2)	8 (2)	0.562
Others	20 (4)	1 (2)	16 (4)	1.000
Non-Caucasian	56 (11)	6 (15)	45 (10)	0.427
Indication for cardiac transplantation				
Coronary artery disease	185 (37)	16 (37)	160 (37)	0.866
Non-ischaemic cardiomyopathy	256 (51)	22 (54)	228 (53)	1.000
Other indications	38 (8)	3 (7)	43 (10)	0.785
Unknown indication	20 (4)	0 (0)	0 (0)	–
Interval between transplantation and study enrollment				
Enrolled prior to transplantation	70 (14)	6 (15)	62 (14)	1.000
< 6 months	367 (7)	29 (71)	323 (75)	0.575
6–12 months	24 (5)	2 (5)	15 (3)	0.652
13–36 months	20 (4)	2 (5)	18 (4)	0.689
37–60 months	2 (0)	0 (0)	2 (0)	1.000
> 60 months	16 (3)	2 (5)	11 (3)	0.314
Cytomegalovirus serology (IgG) status				
Donor and recipient positive	150 (30)	13 (32)	129 (30)	0.859
Donor and recipient negative	78 (16)	5 (12)	68 (16)	0.656
Donor positive and recipient negative	84 (17)	8 (20)	72 (17)	0.663
Donor negative and recipient positive	113 (23)	8 (20)	99 (23)	0.700
Unknown	74 (15)	7 (17)	63 (15)	0.647
Use of ventricular assist device before transplantation	118 (24)	6 (15)	110 (26)	0.133
Induction therapy (any)	384 (77)	34 (83)	295 (68)	0.074
Immunosuppressive therapy, at any time during study				
Cyclosporin A	298 (60)	29 (71)	256 (59)	0.183
Tacrolimus	309 (62)	24 (59)	265 (61)	0.739
Mycophenolate mofetil or mycophenolic acid	421 (84)			
Prednisone	436 (87)	39 (95)	394 (91)	0.561
Left ventricular ejection fraction at first study visit, mean % (standard deviation)	64.1 (9.6)	63.1 (8.7)	64.7 (9.5)	0.301

Values are n (%) unless indicated otherwise.

IgG, immunoglobulin G.

<sup>a</sup>Twenty-seven patients did not have central pathology readings.

first year post-transplantation in the often-cited Kubo article.<sup>17</sup> The lower prevalence of acute cellular rejection in CARGO II results compared with estimates from experiences 20 years earlier may be attributable to more effective contemporary immunosuppression regimens. This prevalence of biopsy-based moderate to severe rejection grades in the current study was consistent with the more recent report of a similar population undergoing biopsies in the first year post-transplantation: biopsy-based acute rejections were

reported to range from 1.9% (35 of 1875 biopsies) to 3.1% (58 of 1854 biopsies).<sup>18</sup>

The current report finds that rates of rejection were similar in the early post-transplantation months (3.2% for months ≥ 2 to 6) compared with the later time post-transplantation (3.2% > Month 6). However, it is possible that the biopsies performed after Month 6 post-transplantation may have been performed more often due to clinical suspicion of rejection, whereas biopsies performed earlier



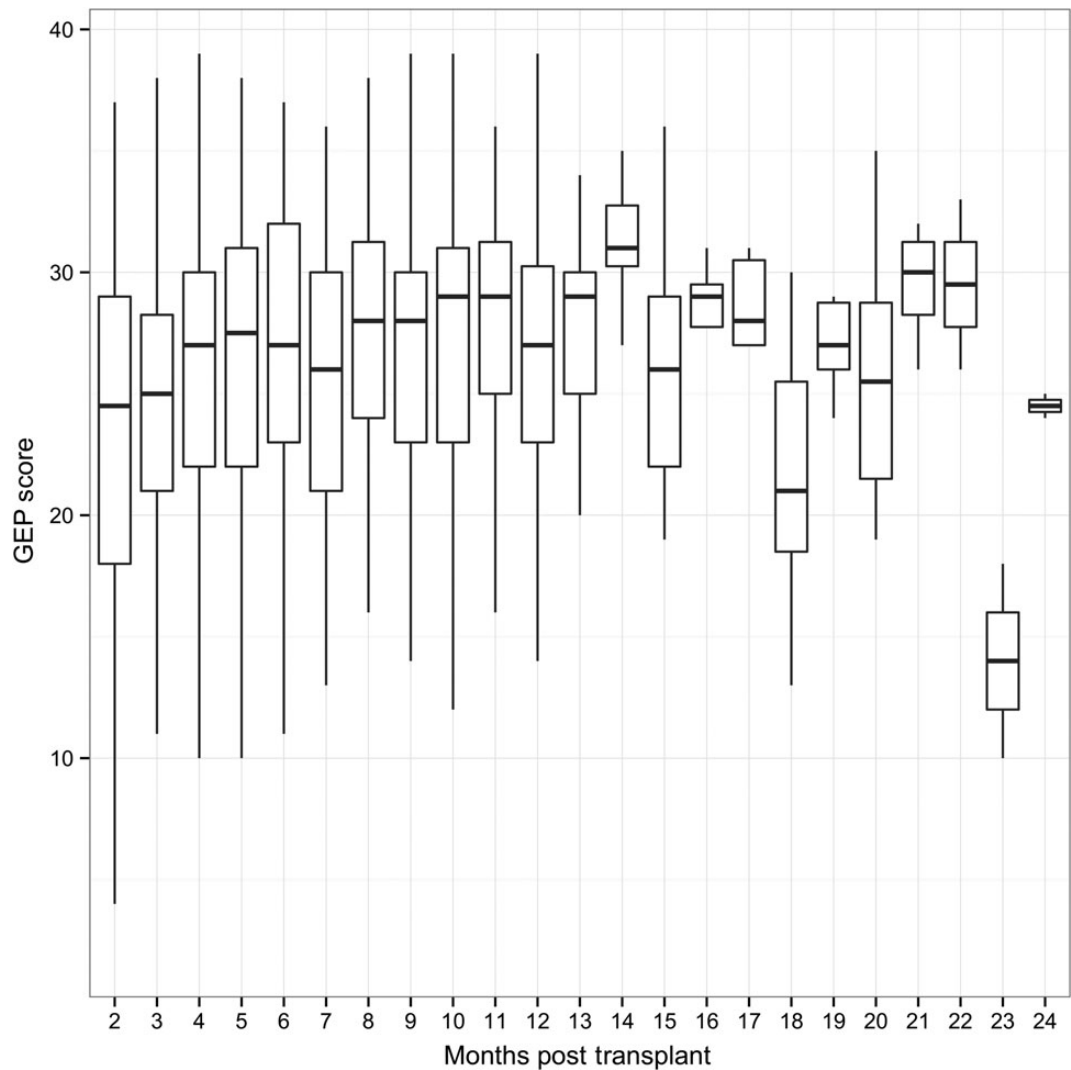
than Month 6 were probably more often conducted for surveillance of asymptomatic patients. This ascertainment bias is discussed further in the limitations section below.

The local biopsy-based rejection cases [grades  $\geq 3A$  (2R)] that were confirmed by the independent panel of central pathologists in the current study served as the key reference standard for

rejection that was used to judge the diagnostic performance of the GEP score. The reason these central pathologist confirmed rejections represent the most important factor in determining the clinical performance of the GEP test is that a biopsy finding of a rejection grade  $\geq 3A$  (2R) generally calls for treatment of the patient, whereas for lower rejection grades or quiescence, there is the option to observe the patient's interval clinical progress at a future visit. All 646 (68.9%) centrally graded biopsies that were by default assigned to an intermediate class (i.e. those biopsies that did not meet criteria to serve as rejection or quiescence reference standards) were included in the computation of the NPV and PPV; these cases were categorized as 'non-rejection'. The relatively large number of biopsies assigned to this intermediate class reflects the high level of discordance between pathologists' readings in this large proportion of biopsies with disparate reader scores of Grade 0, Grade 1A, Grade 1B, and Grade 2. Previous analysis of the CARGO

**Table 2** Agreement between local and central pathology grades for rejection

Local grades	Central grades	
	$\geq 3A$ (2R) rejection	$< 3A$ (2R)
$\geq 3A$ (2R), <i>n/n</i> (%)	46/76 (60.5)	30/76 (39.5)
$< 3A$ (2R), <i>n/n</i> (%)	63/862 (7.3)	799/862 (92.7)



**Figure 2** Gene-expression profiling score by the time post-transplant. The number of gene-expression profiling tests between Month 12 and Month 13 post-transplant ranged from 24 (Month 11) to 128 (Month 3). The number of gene-expression profiling tests between Month 14 and Month 24 post-transplant ranged from 2 (Month 24) to 11 (Month 15).

**Table 3** CARGO II GEP test performance results in the  $\geq 2$ –6 months post-transplantation period

GEP score	Scores below threshold <i>n</i> (%)	PPV (%)	NPV (%)	Sensitivity (%)	Specificity (%)
12	14 (2.9)	1.9	100.0	100.0	3.5
13	21 (4.4)	2.0	100.0	100.0	5.0
14	31 (6.5)	2.0	100.0	100.0	7.1
15	39 (8.1)	2.1	100.0	100.0	9.3
16	50 (10.4)	2.1	100.0	100.0	11.7
17	58 (12.1)	2.1	100.0	100.0	13.3
18	71 (14.8)	2.1	99.4	95.0	15.9
19	83 (17.3)	2.0	99.0	90.0	17.9
20	95 (19.8)	2.0	98.8	87.5	20.2
21	111 (23.1)	2.1	99.0	87.5	23.9
22	133 (27.7)	2.3	99.2	87.5	28.2
23	155 (32.3)	2.1	98.7	77.5	32.7
24	182 (37.9)	2.0	98.3	65.0	38.0
25	201 (41.9)	2.1	98.4	65.0	42.0
26	223 (46.5)	2.1	98.4	60.0	47.2
27	256 (53.3)	2.1	98.3	52.5	53.6
28	295 (61.5)	2.1	98.3	42.5	61.9
29	321 (66.9)	2.1	98.3	37.5	67.3
30	349 (72.7)	2.6	98.4	37.5	73.2
31	373 (77.7)	3.1	98.5	37.5	77.7
32	393 (81.9)	3.1	98.4	30.0	81.9
33	409 (85.2)	3.9	98.5	30.0	85.8
34	425 (88.5)	4.0	98.4	25.0	88.7
35	443 (92.3)	2.3	98.2	10.0	91.9
36	453 (94.4)	3.1	98.2	10.0	94.1
37	460 (95.8)	4.2	98.2	10.0	95.6
38	475 (99.0)	0.0	98.1	0.0	98.9
39	479 (99.8)	0.0	98.1	0.0	99.7

GEP test scores range from 0 to 39.

CARGO, Cardiac Allograft Rejection Gene Expression Observational Study; GEP, gene-expression profiling; NPV, negative predictive value; PPV, positive predictive value.

II data set demonstrated that both within the panel agreement and between the panel and local pathologists agreement decreased with increasing category of severity of rejection grade (from  $\sim 65\%$  to  $\sim 28\%$ ).<sup>10</sup> In that study, an EMB graded  $\geq 3A$  (2R) by one pathologist was unlikely to be equally graded by a given second pathologist. The observed low agreement between pathologists of 28% for grade  $\geq 3A$  (2R) EMB readings indicates the uncertainty of biopsy-based diagnosis of rejection, albeit this method remains the 'gold' standard for assessing the rejection status of a transplanted heart. In the current study, in  $\sim 40\%$  of cases (Table 2), a central pathologist 'downgraded' the local pathologists' grades of  $\geq 3A$  (2R) to a grade  $< 3A$  (2R). The tendency for local pathologists to assign grades of  $\geq 3A$  (2R) more often than central pathologists may be attributable to the local pathologist being influenced by clinical information that is not available to the central reviewers. The discordance of grade assignment by local and central pathologists is also attributable to the diagnosis depending on subjective pattern recognition as well as potential tissue sampling errors and/or the non-specificity of the histology findings.<sup>10,11</sup>

A major reason the ISHLT 2005 rejection grading re-classification combined the several 1990 subcategories of intermediate rejection

grades is that there is uncertainty of the clinical relevance and prognostic implications of these intermediate grades. Most often, a biopsy finding of an intermediate grade will not lead to treatment or other interventional actions: follow-up depends on the clinical context. The GEP test is not designed to correlate with or classify intermediate grades of rejection. However, a sub-analysis of the original CARGO data demonstrated that GEP scores discriminate moderate to severe rejection from mild rejection (Grades 1A and 2). A subgroup of mild rejection cases, defined as Grade 1B according to the 1990 grading system, share a molecular signature more consistent with moderate to severe rejection.<sup>19</sup> The clinical relevance of these intermediate grades of rejection is not part of the current CARGO II study and remains to be defined.

Because the AUC-ROC is easy to understand and it does not rely on specific thresholds, it is commonly used to measure the accuracy of a diagnostic test.<sup>13</sup> Values for the AUC-ROC range from 0.5 (uninformative) to 1.0 (perfect discrimination). An example of AUC-ROC in the cardiovascular area is The Framingham Risk Score, which is based on traditional cardiovascular risk factors.<sup>13</sup> This score yields an AUC-ROC of 0.68, based on age and gender, and goes up

**Table 4** CARGO II GEP test performance results in the >6 months post-transplantation period

GEP score	Scores below threshold n (%)	PPV (%)	NPV (%)	Sensitivity (%)	Specificity (%)
12	6 (1.3)	2.0	100.0	100.0	1.5
13	9 (2.0)	2.0	100.0	100.0	2.1
14	14 (3.1)	2.0	98.5	97.7	3.1
15	19 (4.1)	2.0	98.9	97.7	4.0
16	27 (5.9)	2.1	99.2	97.7	5.9
17	32 (7.0)	2.1	99.3	97.7	7.0
18	43 (9.4)	2.2	99.5	97.7	9.7
19	55 (12.0)	2.2	99.6	97.7	12.1
20	62 (13.5)	2.3	99.7	97.7	14.0
21	79 (17.2)	2.4	99.7	97.7	17.8
22	95 (20.7)	2.5	99.8	97.7	21.7
23	108 (23.6)	2.6	99.8	97.7	25.0
24	130 (28.4)	2.8	99.8	97.7	30.3
25	146 (31.9)	2.8	99.6	93.2	33.3
26	165 (36.0)	3.0	99.6	93.2	38.7
27	198 (43.2)	3.2	99.4	86.4	46.5
28	220 (48.0)	3.1	99.1	77.3	51.2
29	251 (54.8)	3.4	99.0	72.7	57.4
30	294 (64.2)	3.8	98.9	63.6	67.6
31	331 (72.3)	4.4	98.8	54.5	76.1
32	357 (77.9)	4.6	98.6	45.5	80.6
33	378 (82.5)	3.5	98.3	27.3	84.5
34	399 (87.1)	4.3	98.3	25.0	88.8
35	418 (91.3)	2.8	98.1	11.4	92.0
36	439 (95.9)	4.7	98.1	9.1	96.2
37	450 (98.3)	0.0	98.0	0.0	98.2
38	454 (99.1)	0.0	98.0	0.0	99.4
39	455 (99.3)	0.0	98.0	0.0	99.8

GEP test scores range from 0 to 39.  
CARGO, Cardiac Allograft Rejection Gene Expression Observational Study; GEP, gene-expression profiling; NPV, negative predictive value; PPV, positive predictive value.

**Table 5** True positive, false positive, true negative, and false negative rates based on GEP threshold score  $\geq 34$

	$\geq 3A$ (2R)	$< 3A$ (2R)
GEP score	$\geq 2-6$ months post-transplant	
$\geq 34$ , n/n (%)	5/55 (9.1) TP	50/55 (90.9) FP
$< 34$ , n/n (%)	17/425 (4.0) FN	408/425 (96.0) TN
GEP n	$> 6$ months post-transplant	
$\geq 34$ , n/n (%)	6/59 (10.2) TP	53/59 (89.8) FP
$< 34$ , n/n (%)	18/399 (4.5) FN	381/399 (95.5) TN

Example: 5/55, 5 = number of cases with GEP score  $\geq 34$  and  $\geq 3A$  (2R) rejection; 55 = sum of  $\geq 3A$  (2R) and  $< 3A$  (2R) rejections with GEP score of  $\geq 34$ .  
GEP, gene-expression profiling; FN, false negative; FP, false positive; TN, true negative, TP, true positive.

from the  $\geq 2-6$  months and  $> 6$  month time windows, respectively, were comparable to the CARGO results (see Supplementary material online, Appendix S1).<sup>6</sup> The limited sensitivity and PPV of the GEP score for a histology diagnosis of acute cellular rejection are similar to the sensitivity and PPV reported between expert pathologists.<sup>10,11</sup> Observed concordance between local and collaborating study pathologists for EMB scoring was 28% for grade  $\geq 3A$  (2R) in the CARGO II study<sup>10</sup> and 40% for grade  $\geq 3A$  (2R) using 1990 ISHLT grading in the CARGO Study.<sup>11</sup> This circumstance of a low concordance of rejection grades between expert pathologists has also been reported with interpretation of biopsies from transplanted kidneys.<sup>20,21</sup> Kidney transplant researchers who are seeking to develop molecular signatures that are correlated to histology reference standards of rejection have observed a problem that applies also to heart transplant: the 'reference standard-related bias'. The new test, whether based on a molecular classifier or the opinion of a second pathologist, will not agree perfectly with the existing reference standard (the opinion of a different pathologist), when there is a high degree of subjectivity involved in the assignment of the reference standard. 'Noise' and the associated range of uncertainty within heart (or other transplanted organ such as kidney) tissue

to 0.75 when other risk factors are added. The CARGO II GEP AUC-ROC of 0.70 (95% confidence interval 0.67 to 0.73) and 0.69 (95% confidence interval 0.66 to 0.72) for patient samples



**Table 6** Univariate and multivariate analyses for variables predictive of acute cellular rejection  $\geq 3A$  (2R)

Characteristics	Univariate		Multivariate	
	Coefficient <sup>a</sup>	P-values	Coefficient <sup>a</sup>	P-values
Gene-expression profiling score (incremental unit increase by one)	0.08	0.004	0.08	0.004
Age, $\geq 54$	-0.02	0.044	-0.03	0.008
Gender, male	0.09	0.811	0.25	0.549
Primary indication for cardiac transplantation, ischaemic	-0.02	0.961	-0.28	0.446
Cytomegalovirus serology status, donor or recipient positive	0.42	0.332	0.39	0.382
Left ventricular assist device use, yes	-0.59	0.198	-0.87	0.069

<sup>a</sup>Positive coefficient indicates a probability of acute cellular rejection  $\geq 3A$  (2R) increases with specified variable setting; negative coefficient indicates decreased risk of acute cellular rejection  $\geq 3A$  (2R).

**Table 7** Univariate analyses of clinical variables for significance of gene-expression profiling score in estimating risks of  $\geq 3A$  (2R) acute cellular rejection

Covariate	$\geq 3A$ (2R) rejection (n)	Non-rejection (n)	P-values
Age			
$\geq 54$ years	13	447	0.074
$< 54$ years	33	445	0.021
Gender			
Male	37	706	0.004
Female	9	186	0.467
Primary indication for cardiac transplantation			
Ischaemic	17	326	0.003
Other	29	566	0.164
Cytomegalovirus serology status, donor or recipient positive			0.39
Donor or recipient positive	39	701	0.020
Donor or recipient negative or unknown	7	191	0.107
Left ventricular assist device use			
Yes	6	189	0.153
No	40	703	0.011

Reported P-values represent significance of gene-expression profiling score in estimating risks of  $\geq 3A$  (2R) acute cellular rejection within subsets of patients defined by each covariate.

histology is intrinsic, due to a variety of causes as summarized above.<sup>9,10,21</sup> To transcend this limitation of histology, it may be useful to correlate the performance of GEP use to alternative clinical endpoints such as organ function; this approach is outside the scope of the current paper, but the potential of such approaches has been reported.<sup>22,23</sup>

Although CARGO II, like CARGO, was a landmark study in cumulative number of patients and surveillance samples/visits collected, there was a low cumulative incidence of observed rejections. The relatively small number of rejection episodes and the spread of these cases over a wide post-transplantation time range likely constrain the GEP test performance characteristics (i.e. failure of GEP test to achieve statistical significance in some of our subgroup

analyses could be attributed to imbalance in the dataset and to the very low sample size for some covariates, Table 7). The GEP score was an equally significant predictor of acute cellular rejection ( $\geq 2R$ ) in both univariate and multivariate models (Table 6). Nevertheless, the discovery of asymptomatic rejection may be enhanced if a GEP score exceeding threshold (i.e. 34) is used to lead to performance of a biopsy, since the GEP score PPV in this circumstance may be approximately two-fold higher than the prevalence of rejection in the surveillance population.

The AUC-ROC was analysed in two time windows, because the mean GEP score in a large, unselected, patient population tends to gradually rise from Month 2 to Month 6; this upward trend is less pronounced from Month 7 to 12; the mean GEP score reaches a

plateau and remains stable from Years 1 to 5 post-transplantation.<sup>7</sup> Findings from this study confirmed the GEP score trajectory between Months 2 and 12 post-transplant. After 12 months, the CARGO II mean GEP scores are much more variable and lower than observed in the larger database. This discrepancy may be explained by the small number of samples available from the current study after 12 months (Figure 2) and patient selection differences between the studies. The early time dependency of average GEP scores and acquired practical experience has led many centres to use 34 as the selected threshold to rule out rejection for patients who are >6–12 months post-transplantation, and to consider a lower threshold (e.g. 30) for patients who are 2–6 months post-transplantation.<sup>24</sup>

The GEP test is 'intended to aid in the identification of heart transplant recipients with stable graft function who have a low probability of moderate/severe ACR at the time of testing in conjunction with standard clinical assessment' (FDA cleared intended use statement). Specific threshold values are not designated in the approved labeling; rather, the clinician may choose how to select the test score threshold and make an associated clinical interpretation based on his/her experience with the test and the overall, comprehensive clinical context of the patient under evaluation. Thus, the complete listing of NPVs and PPVs for each GEP score from 0 to 39 is provided for the clinician to refer to in considering how a specific test score may aid in 'ruling out' ACR. The threshold of  $\geq 34$  was used in the IMAGE outcomes trial: for patients at least 6 months post-transplant, a score of  $< 34$  was assumed in IMAGE to be equivalent to a biopsy diagnosis of no rejection. In addition, useful clinical information may also be obtained by considering findings from a series of GEP scores in a patient, since each patient may serve as his/her own control of a 'normal' GEP score: within patient variability of GEP scores (after 1 year post-transplantation) may be less than that between patients.<sup>22</sup>

As previously characterized in the CARGO study,<sup>6</sup> the current CARGO II results confirmed that the high NPV associated with GEP scores below a nominal threshold may be reasonably used in lieu of surveillance biopsies to rule out rejection. The IMAGE<sup>3</sup> and E-IMAGE<sup>24</sup> studies have demonstrated that the number of surveillance biopsies was safely reduced while the clinical outcomes of patients managed with GEP surveillance were non-inferior to outcomes in patients managed with the biopsy strategy.

The results of this study must be interpreted in the context of some limitations. The study excluded GEP data and biopsies from visits earlier than 55 days post-transplantation, so the analyses do not inform us about GEP performance or rejection incidence in the first 54 days post-transplantation. Also, because patients treated for rejection within the prior 30 days were excluded, we may have underestimated the incidence of rejection that might be found in the entire population. Additionally, in this study, not all GEP tests had paired biopsy results. Because of the very small number and proportion of clinically suspected, biopsy confirmed, rejections of grade  $\geq 3A$  (2R), rejections discovered during routine surveillance and rejections confirmed with biopsies performed for clinical suspicion were pooled together in the analyses. These reference rejection cases, a combination of clinically silent and clinically suspected rejection, may not have identical GEP signatures. Although the GEP test is intended for use in a surveillance setting, the discovery of subclinical rejection is dependent on the frequency of surveillance that varied

at the participating centres. Additionally, the histology grading is an imperfect reference standard for diagnosis of rejection, although it remains the clinical standard of care for diagnosis of rejection in heart transplantation.

The central pathology feature of this study does not reflect practical clinical care: no clinician relies on obtaining additional histopathology reports from 'blinded' independent pathologists to decide on treatment of a patient. We utilized the central biopsy reading in this study to enhance the accuracy of diagnostic classification of the reference cases. Finally, the question remains whether GEP scores can be potentially used to optimize a post-transplant immunosuppression regimen. This study was not designed to address that question. Future studies should be conducted to evaluate the clinical utility of the GEP test to guide immunosuppression medication intensity to improve long-term outcomes in transplant recipients.

## Conclusion

The GEP test performance characteristics in this CARGO II study are consistent with prior CARGO findings. The low prevalence of ACR contributes to the high NPV and limited PPV of GEP testing. Taken together with clinical trial evidence that equivalent clinical outcomes may be attained in patients managed with GEP compared with patients managed with biopsies for acute cellular rejection surveillance, the CARGO II results further validate that GEP testing may be used as part of a non-invasive surveillance strategy to rule out acute cellular rejection in heart transplant recipients. The choice of threshold score for practical use of GEP testing to rule out rejection should consider overall clinical assessment of the patient's baseline risk for rejection.

## Supplementary material

Supplementary material is available at *European Heart Journal* online.

## Authors' contributions

B.E. performed statistical analysis. J.P.Y. handled funding and supervision. M.G.C.-L., J.S., U.S., A.Z., P.M., C.B., H.R., J.P., M.Z., R.F., D.H., M.C.D., P.L., and J.V. acquired the data. M.G.C.-L., J.S., A.Z., P.M., and J.V. conceived and designed the research. J.V., M.G.C.-L., J.S., and J.P.Y. drafted the manuscript. P.M., M.C., J.P.Y., and H.R. made critical revision of the manuscript for key intellectual content.

## Acknowledgements

David Hiller, PhD, provided statistical support and Helen Nicely, PhD, provided editorial assistance.

## Funding

This work was supported by CareDx Inc. (ClinicalTrials.gov Identifier: NCT00761787). Funding to pay the Open Access publication charges for this article was provided by CareDx, Inc.

**Conflict of interest:** M.C.D. serves on the advisory board of CareDx. J.S., U.S., P.M., A.Z., and M.C.D. received study support from CareDx. J.P.Y. is an employee owning equity in CareDx. B.E. has been a statistical

consultant to CareDx. No other potential conflict of interest relevant to this article was reported.

## References

- Deng MC, Eisen HJ, Mehra MR, Billingham M, Marboe CC, Berry G, Kobashigawa J, Johnson FL, Starling RC, Murali S, Pauly DF, Baron H, Wohlgemuth JG, Woodward RN, Klingler TM, Walther D, Lal PG, Rosenberg S, Hunt S, CARGO Investigators. Noninvasive discrimination of rejection in cardiac allograft recipients using gene expression profiling. *Am J Transplant* 2006;**6**:150–160.
- Starling RC, Pham M, Valantine H, Miller L, Eisen H, Rodriguez ER, Taylor DO, Yamani MH, Kobashigawa J, McCurry K, Marboe C, Mehra MR, Zuckerman A, Deng MC, Working Group on Molecular Testing in Cardiac Transplantation. Molecular testing in the management of cardiac transplant recipients: initial clinical experience. *J Heart Lung Transplant* 2006;**25**:1389–1395.
- Pham MX, Teuteberg JJ, Kfoury AG, Starling RC, Deng MC, Cappola TP, Kao A, Anderson AS, Cotts WG, Ewald GA, Baran DA, Bogaev RC, Elashoff B, Baron H, Yee J, Valantine HA, IMAGE Study Group. Gene expression profiling for rejection surveillance after cardiac transplantation. *N Engl J Med* 2010;**362**:1890–1900.
- Costanzo MR, Dipchand A, Starling R, Anderson A, Chan M, Desai S, Fedson S, Fisher P, Gonzales-Stawinski G, Martinelli L, McGiffin D, Smith J, Taylor D, Meiser B, Webber S, Baran D, Carboni M, Dengler T, Feldman D, Frigerio M, Kfoury A, Kim D, Kobashigawa J, Shullo M, Stehlik J, Teuteberg J, Uber P, Zuckermann A, Hunt S, Burch M, Bhat G, Canter C, Chinnock R, Crespo-Leiro M, Delgado R, Dobbels F, Grady K, Kao W, Lamour J, Parry G, Patel J, Pini D, Towbin J, Wolfel G, Delgado D, Eisen H, Goldberg L, Hosenpud J, Johnson M, Keogh A, Lewis C, O'Connell J, Rogers J, Ross H, Russell S, Vanhaecke J, International Society of Heart and Lung Transplantation Guidelines. The International Society of Heart and Lung Transplantation Guidelines for the care of heart transplant recipients. *J Heart Lung Transplant* 2010;**29**:914–956.
- Hresko A, Haga SB. Insurance coverage for personalized medicine. *J Pers Med* 2012;**2**:201–216.
- CareDx (XDx) GEP Laboratory Services Guide, LQ10004R2. [http://www.allomap.com/wp-content/uploads/2014/04/XDx\\_LabServicesGuideFINAL\\_3electronicversion.pdf](http://www.allomap.com/wp-content/uploads/2014/04/XDx_LabServicesGuideFINAL_3electronicversion.pdf) (02 August 2015).
- Austin BA, Arnold PJ, Kao A. The impact of time post cardiac transplant on gene expression profile scores, an analysis of 32,043 Tests. *J Cardiovasc Dis Diag* 2013;**1**:114.
- Billingham ME, Cary NR, Hammond ME, Kemnitz J, Marboe C, McCallister HA, Snovar DC, Winters GL, Zerbe A. A working formulation for the standardization of nomenclature in the diagnosis of heart and lung rejection. *J Heart Transplant* 1990;**9**:587–593.
- Stewart S, Winters GL, Fishbein MC, Tazelaar HD, Kobashigawa J, Abrams J, Andersen CB, Angelini A, Berry GJ, Burke MM, Demetris AJ, Hammond E, Itescu S, Marboe CC, McManus B, Reed EF, Reinsmoen NL, Rodriguez ER, Rose AG, Rose M, Suciu-Foca N, Zeevi A, Billingham ME. Revision of the 1996 working formulation for the standardization of nomenclature in the diagnosis of heart rejection. *J Heart Lung Transplant* 2005;**24**:1710–1720.
- Crespo-Leiro MG, Zuckermann A, Bara C, Mohacsi P, Schulz U, Boyle A, Ross HJ, Parameshwar J, Zakliczynski M, Fiocchi R, Stypmann J, Hoefer D, Lehmkühl H, Deng MC, LePrince P, Berry G, Marboe CC, Stewart S, Tazelaar HD, Baron HM, Coleman IC, Vanhaecke J. Concordance among pathologists in the second cardiac allograft rejection gene expression observational study (CARGO II). *Transplantation* 2012;**94**:1172–1177.
- Marboe CC, Billingham M, Eisen H, Deng MC, Baron H, Mehra M, Hunt S, Wohlgemuth J, Mahmood I, Prentice J, Berry G. Nodular endocardial infiltrates (Quilty lesions) cause significant variability in diagnosis of ISHLT Grade 2 and 3A rejection in cardiac allograft recipients. *J Heart Lung Transplant* 2005;**24**(7 Suppl): 219–S226.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.
- Wang TJ. Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. *Circulation* 2011;**123**:551–565.
- Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Acad Radiol* 2005;**12**:1198–1204.
- Emir B, Wieand S, Su JQ, Cha S. Analysis of repeated markers used to predict progression of cancer. *Stat Med* 1998;**17**:2563–2578.
- Stehlik J, Edwards LB, Kucheryavaya AY, Benden C, Christie JD, Dipchand AI, Dobbels F, Kirk R, Rahmel AO, Hertz MI, International Society of Heart and Lung Transplantation. The Registry of the International Society for Heart and Lung Transplantation: 29th official adult heart transplant report-2012. *J Heart Lung Transplant* 2012;**31**:1052–1064.
- Kubo SH, Naftel DC, Mills RM Jr, O'Donnell J, Rodeheffer RJ, Cintron GB, Kenzora JL, Bourge RC, Kirklin JK. Risk factors for late recurring rejection after heart transplantation: a multiinstitutional, multivariable analysis. *J Heart Lung Transplant* 1995;**14**:409–418.
- Hamour IM, Burke MM, Bell AD, Panicker MG, Banerjee R, Banner NR. Limited utility of endomyocardial biopsy in the first year after heart transplantation. *Transplantation* 2008;**85**:969–974.
- Bernstein D, Williams GE, Eisen H, Mital S, Wohlgemuth JG, Klingler TM, Fang KC, Deng MC, Kobashigawa J. Gene expression profiling distinguishes a molecular signature for grade 1B mild acute cellular rejection in cardiac allograft recipients. *J Heart Lung Transplant* 2007;**26**:1270–1280.
- Reeve J, Sellarés J, Mengel M, Sis B, Skene A, Hidalgo L, de Freitas DG, Famulski KS, Halloran PF. Molecular diagnosis of T cell-mediated rejection in human kidney transplant biopsies. *Am J Transplant* 2013;**13**:645–655.
- Halloran PF, Pereira AB, Chang J, Matas A, Picton M, De Freitas D, Bromberg J, Serón D, Sellarés J, Einecke G, Reeve J. Potential impact of microarray diagnosis of T cell-mediated rejection in kidney transplants: the INTERCOM study. *Am J Transplant* 2013;**13**:2325–2326.
- Deng MC, Elashoff B, Pham MX, Teuteberg JJ, Kfoury AG, Starling RC, Cappola TP, Kao A, Anderson AS, Cotts WG, Ewald GA, Baran DA, Bogaev RC, Shahzad K, Hiller D, Yee J, Valantine HA, IMAGE Study Group. Utility of gene expression profiling score variability to predict clinical events in heart transplant recipients. *Transplantation* 2014;**97**:708–714.
- Crespo-Leiro MG, Stypmann J, Zuckermann A, Bara C, Ross H, Parameshwar J, Zakliczynski M, Fiocchi R, Hoefer D, Hiemann N, LePrince P, Deng MC, Hiller D, Yee JP, Vanhaecke J. Utility of gene expression profiling test (GEP) score instability to predict future clinical outcomes in heart transplant: results from the CARGO 2 European-based multicenter trial. *J Heart Lung Transplant* 2013;**32**:S113–S114.
- Kobashigawa J, Patel J, Azarbal B, Kittleson M, Chang D, Czer L, Daun T, Luu M, Trento A, Cheng R, Esmailian F. Randomized pilot trial of gene expression profiling versus heart biopsy in the first year after heart transplant: early invasive monitoring attenuation through gene expression trial. *Circ Heart Fail* 2015;**8**:557–564.